

Efficient Capacity Computation and Power Optimization for Relay Networks

Farzad Parvaresh and Raúl Etkin

Abstract—The capacity or approximations to capacity of various single-source single-destination relay network models has been characterized in terms of the cut-set upper bound. In principle, a direct computation of this bound requires evaluating the cut capacity over exponentially many cuts. We show that the minimum cut capacity of a relay network under some special assumptions can be cast as a minimization of a submodular function, and as a result, can be computed efficiently. We use this result to show that the capacity, or an approximation to the capacity within a constant gap for the Gaussian, wireless erasure, and Avestimehr-Diggavi-Tse deterministic relay network models can be computed in polynomial time. We present some empirical results showing that computing constant-gap approximations to the capacity of Gaussian relay networks with around 300 nodes can be done in order of minutes.

For Gaussian networks, cut-set capacities are also functions of the powers assigned to the nodes. We consider a family of power optimization problems and show that they can be solved in polynomial time. In particular, we show that the minimization of the sum of powers assigned to the nodes subject to a minimum rate constraint (measured in terms of cut-set bounds) can be computed in polynomial time. We propose an heuristic algorithm to solve this problem and measure its performance through simulations on random Gaussian networks. We observe that in the optimal allocations most of the power is assigned to a small subset of relays, which suggests that network simplification may be possible without excessive performance degradation.

Index Terms—capacity, network simplification, power allocation, relay networks, submodular optimization.

I. INTRODUCTION

Relay networks, where one or more source nodes send information to one or more destination nodes with the help of intermediate nodes acting as relays, are often used to model communication in wireless sensor networks. In sensor networks, sensor nodes have limited power sources and often require multi-hop communication with the help of intermediate nodes to reach the data aggregation centers. To guide the design of these networks it is of interest to characterize fundamental communication limits such as the capacity, which represents the maximum reliable communication rate.

Various communication models for relay networks capture in an abstract setting different aspects of practical systems. The wireless erasure network model of [8] captures the effect of packet losses in the wireless setting. The deterministic network model of Avestimehr, Diggavi and Tse (ADT) [4] incorporates broadcast and interference and can be used to gain insights about communication in more complex models that incorporate noise. Among these, of special importance

is the Gaussian relay network, which models power limited transmitters and received signals corrupted by additive white Gaussian noise.

While the capacity of some network models (e.g. wireless erasure and ADT) is well characterized, the capacity of the Gaussian relay network, even in its simplest form with one transmitter, one relay, and one receiver, is in general unknown. The best known capacity upper bound is the so-called *cut-set bound*. A cut Ω of a network can be considered as a subset of nodes which includes the source node and excludes the destination node. For this cut, the capacity $F(\Omega)$ is defined as the maximum rate that information can be transferred from the nodes in Ω to the nodes that are not in Ω conditioned on the fact the information on Ω^c (the nodes that are not in Ω) is known. The cut-set upper bound is the *minimum* cut capacity over all the possible cuts.

In the Gaussian setting, there are several capacity lower bounds based on different communication schemes, such as amplify-and-forward, decode-and-forward, compress-and-forward, quantize-and-forward, etc. [5], [7], [18]. Recently, Avestimehr, et al. [2] made significant progress in the capacity characterization of Gaussian relay networks by showing that a quantization and coding communication scheme can achieve a communication rate within a constant gap of the cut-set upper bound, where the gap only depends on the number of nodes in the network (i.e. it is independent of the channel gains and power levels). However, the evaluation of the achievable communication rate, which is necessary to implement the scheme, requires the computation of the cut-set bound for the network. Assuming that for a given cut the cut capacity is easy to compute, finding the cut-set upper bound can be a challenging problem. For a network with n relays there are 2^n different cuts and a greedy algorithm needs exponential time in the number of relays to compute the cut-set bound.

In this work we show that the achievable rate of the scheme of [2] for the Gaussian relay network can be computed in polynomial time, and as a result, can be computed efficiently. This result is obtained by showing that the cut capacity of a fairly large class of networks under the assumption of independent encoding at the nodes in Ω is a submodular function. For the special case of layered relay networks, [27] showed the equivalent of our submodularity result simultaneously with our conference version of this paper [25]. Submodularity properties of conditional entropy (in terms of which cut-capacities are expressed) have also been used in [2], [3] to bound the cut-capacity of a network in terms of the cut-capacity of the corresponding unfolded graph¹.

F. Parvaresh and R. Etkin are with Hewlett-Packard Laboratories, Palo Alto, CA 94304, USA. (emails: {parvaresh, raul.etkin}@hp.com).

Preliminary conference versions of the results in this paper appeared in [25].

¹Please, refer to [2] for the definition of an unfolded graph.

Existing results on minimization of submodular functions provide algorithms with polynomial running time $O(n^5\alpha + n^6)$, where α is the time that it takes to compute $F(\Omega)$ and n is the number of nodes in the network [20]. In addition, there exist possibly faster algorithms without polynomial time performance guarantees based on Wolfe's minimization norm algorithm [12]. In Section VI, by simulations, we show that the cut-set bound for a Gaussian relay network with around 300 nodes can be computed on a laptop computer in about a minute using a Matlab package for submodular minimization provided in [19].

Our results, extend and generalize previous results for the ADT model. This model can be seen as a high signal-to-noise-ratio (SNR) approximation of the Gaussian model, incorporating the effects of broadcasting and superposition of signals while de-emphasizing the effects of noise. Amduruz et al. [1] showed that the cut-set bound for a *layered*² ADT model can be computed efficiently. They have extended graph flow algorithms such as Ford-Fulkerson's in a nontrivial way to find the maximum possible *linearly independent* (LI) paths in the network. They showed that the capacity of the network is equal to the maximum number of (LI) paths and can be computed in time $O(M \cdot |E| \cdot C^5)$, where M is the maximum number of nodes per layer, $|E|$ is the total number of edges and C is the capacity of the network. Moreover, they showed that the capacity can be achieved by using a very simple one-bit processing at the relay nodes. Later Goemans et al. [13] showed that the deterministic model is a special case of a flow model based on linking systems, a combinatorial structure with a tight connection to matroids. As a by-product, they obtained the submodularity of the cut capacity for layered ADT networks. Using this observation they provided various algorithms related to matroid theory to compute the cut capacity of the layered deterministic model based on finding intersection or partition of matroids. These results led to faster algorithms to compute the capacity of large layered ADT networks. In addition, there has been other extensions on improving the running time of the current algorithms for computing the capacity of ADT networks [9], [10], [22], [23].

In addition to showing that the capacity within a constant gap of the Gaussian relay network can be computed in polynomial time, our results allow us to compute in polynomial time the capacity of the wireless erasure network. Furthermore, we provide a simple proof for the computability in polynomial time of the capacity of the layered and non-layered ADT networks.

Building on the submodularity of the cut-capacity for independent encoding at the nodes, we show that, in the Gaussian setting, it is possible to efficiently optimize the power allocated to the source and relay nodes. We consider two power optimization problems: (i) minimize the total power satisfying a minimum source-destination data rate constraint and power constraints at each node; (ii) maximize the source-destination data rate satisfying total and individual power

constraints at the nodes. Since the capacity of the Gaussian relay network is approximately given by the cut-set upper bound with independent encoding at the nodes, we use this cut-set bound to characterize data rate in the optimization problems. We show that these optimization problems can be solved in polynomial time and use simulations to get insights about some properties of the optimal power allocations for networks of various sizes. We observe that optimal power allocations assign most of the power to a small subset of nodes and that setting the power to zero in the remaining nodes (i.e. removing these nodes from the network) often results in a small rate loss. Nazaroglu, et al. showed in [26] that for the special case of the N -relay Gaussian diamond network a fraction $k/(k+1)$ of the total capacity can be approximately achieved by using only k of the total N relays. This suggests that the diamond network can be significantly simplified by tolerating a small performance loss. Our results provide a numerical counterpart to the fundamental performance bounds derived in [26] and suggest that network simplification may also be possible in more general Gaussian relay networks.

We obtain these results by considering a general framework to compute the cut-set bound. We assign transmit signal random variable X_i to node $i \in \mathcal{V}$ and we assume the probability distribution over the signals X_1, X_2, \dots, X_n to be independent, i.e. $p(X_1, X_2, \dots, X_n) = p_1(X_1)p_2(X_2) \cdots p_n(X_n)$. We also assign received signal random variables Y_i 's to each node. The network is defined by the transition probability function $f(Y_1, Y_2, \dots, Y_n | X_1, X_2, \dots, X_n)$. We further assume that the transition probability function is of the form $f_1(Y_1 | X_1, \dots, X_n) \cdots f_n(Y_n | X_1, \dots, X_n)$, meaning that the received signals are independent conditioned on the transmitted signals in the network. For such networks we show that $F(\Omega) = I(\mathbf{Y}_{\Omega^c}; \mathbf{X}_{\Omega} | \mathbf{X}_{\Omega^c})$ ³ is submodular with respect to Ω . Later we show that for ADT networks, the Gaussian relay network and the wireless erasure network, we can find $p_1(X_1) \cdots p_n(X_n)$ such that $\min_{\Omega} F(\Omega)$ becomes equal to the capacity or the capacity within a constant gap. In other words, the min-cut problem for these networks can be cast as a minimization of a submodular function.

The paper is organized as follows. In Section III we show that for specific type of networks the cut value, $F(\Omega)$, is a submodular function. We then show in Section IV that for many wireless network models such as the ADT deterministic network, Gaussian relay network and wireless erasure network the capacity or an approximation to the capacity can be cast as a minimization of $F(\Omega)$. In Section V we study two power optimization problems and show that they can be solved efficiently. Finally, in Section VI we describe results related to solving optimization problems involving submodular functions and perform power optimization in various randomly generated networks of different sizes. We start by introducing the notation used in the rest of the paper.

II. NOTATION

Let \mathcal{V} denote the set of nodes in the network and $|\mathcal{V}|$ its cardinality. For any subset A of nodes we denote by $\mathcal{V} \setminus A$

²In a layered network, the nodes in one layer are only connected to the nodes in the next adjacent layer. In particular, there is no direct connection from source to destination.

³See Section II for a definition of the notation \mathbf{X}_{Ω} , \mathbf{Y}_{Ω^c} , etc.

or A^c the set of nodes in \mathcal{V} that are not in A . We assume $\mathcal{V} \setminus \mathcal{A} \cup \mathcal{B} = \mathcal{V} \setminus (\mathcal{A} \cup \mathcal{B})$. A cut Ω is defined as a subset of nodes in \mathcal{V} . A cut splits the nodes in the network into two groups, the nodes that are in Ω and the ones that belong to $\mathcal{V} \setminus \Omega$. Random variables are shown in capital letters such as X_i and Y_i . We use boldface letter for vectors, e.g. \mathbf{x} is a constant vector and \mathbf{X} is a random vector. We use \mathbf{X}_Ω to denote $(X_{v_1}, X_{v_2}, \dots, X_{v_{|\Omega|}})$ with $v_i \in \Omega$. The function $I(X; Y|Z)$ is the mutual information between random variables X and Y conditioned on random variable Z . With a slight abuse of notation we use $H(X)$ to denote either the entropy or differential entropy of the discrete or continuous random variable X [6]. By \mathbb{F}_p we denote a finite field with p elements. Finally, all the $\log(\cdot)$ functions are in base two.

III. SUBMODULARITY OF CUT-SET FUNCTION

Submodularity arises in many combinatorial optimization problems and large body of research has been developed on minimizing or maximizing submodular functions under various constraints.

A submodular function $f : 2^\mathcal{V} \rightarrow \mathbb{R}$ is defined as a function over subsets of \mathcal{V} with *diminishing marginal returns*, i.e. if $A, B \subseteq \mathcal{V}$ with $A \subseteq B$ and any $v \in \mathcal{V} \setminus B$,

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B).$$

The theorem below establishes the submodularity of the cut capacity function of a general relay network under some special assumptions. This theorem will be used in Section IV to prove that the capacity or an approximation to the capacity of various specific relay network models can be computed by minimizing a submodular function.

Theorem 1. *Consider a network consisting of nodes in \mathcal{V} . Each node sends a message $X_i, i \in \mathcal{V}$ and receives $Y_i, i \in \mathcal{V}$. If the messages are independent $p(X_1, X_2, \dots, X_{|\mathcal{V}|}) = p_1(X_1)p_2(X_2) \dots p_{|\mathcal{V}|}(X_{|\mathcal{V}|})$ and conditioned on the sent messages the received messages are independent, then the function*

$$F(A) = I(\mathbf{X}_A; \mathbf{Y}_{\mathcal{V} \setminus A} | \mathbf{X}_{\mathcal{V} \setminus A}), \quad A \subseteq \mathcal{V}$$

is submodular.

Proof: To show that $F(A)$ is submodular we show that $F(A \cup a) - F(A)$ is monotonically non-increasing in A for $a \notin A$.

$$\begin{aligned} F(A \cup a) &= I(\mathbf{X}_{A \cup a}; \mathbf{Y}_{\mathcal{V} \setminus A \cup a} | \mathbf{X}_{\mathcal{V} \setminus A \cup a}) \\ &\stackrel{(a)}{=} H(\mathbf{X}_{A \cup a} | \mathbf{X}_{\mathcal{V} \setminus A \cup a}) - H(\mathbf{X}_{A \cup a} | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A \cup a}) \\ &\stackrel{(b)}{=} H(\mathbf{X}_A) + H(X_a | \mathbf{X}_A) - H(X_a | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A \cup a}) \\ &\quad - H(\mathbf{X}_A | X_a, \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A \cup a}) \\ &= H(\mathbf{X}_A) + H(X_a | \mathbf{X}_A) - H(X_a | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A \cup a}) \\ &\quad - H(\mathbf{X}_A | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A}) \end{aligned}$$

where (a) is the definition of mutual information and (b) is from the chain rule for the entropy function. Therefore,

$$\begin{aligned} F(A \cup a) - F(A) &= H(X_a | \mathbf{X}_A) - H(X_a | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A \cup a}) \end{aligned}$$

$$\begin{aligned} &- H(\mathbf{X}_A | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A}) \\ &+ H(\mathbf{X}_A | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, Y_a, \mathbf{X}_{\mathcal{V} \setminus A}) \\ &= H(X_a | \mathbf{X}_A) - H(X_a | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A \cup a}) \\ &\quad - I(\mathbf{X}_A; Y_a | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A}) \\ &= H(X_a | \mathbf{X}_A) - H(X_a | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A \cup a}) \\ &\quad - H(Y_a | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A}) \\ &\quad + H(Y_a | \mathbf{X}_A, \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A}) \\ &= \underbrace{H(X_a | \mathbf{X}_A)}_{\text{non-increasing in } A} - \underbrace{H(X_a | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A})}_{\text{nondecreasing in } A} \\ &\quad - \underbrace{H(Y_a | \mathbf{Y}_{\mathcal{V} \setminus A \cup a}, \mathbf{X}_{\mathcal{V} \setminus A})}_{\text{nondecreasing in } A} + H(Y_a | \mathbf{X}_{\mathcal{V}}) \end{aligned}$$

where the last equality follows because Y_a is independent of $\mathbf{Y}_{\mathcal{V} \setminus A \cup a}$ conditioned on $\mathbf{X}_{\mathcal{V}}$. So, $F(A \cup a) - F(A)$ is non-increasing in A and thus $F(A)$ is submodular. ■

In the following example we show that if the signals at the nodes are correlated then $F(A)$ is not necessarily a submodular function.

Example. Consider a symmetric Gaussian diamond network with two relays such that the channel gains from source to relays are equal to one and from relays to destination are equal to three. Letting $\mathbf{X}_s, \mathbf{X}_{r_1}$, and \mathbf{X}_{r_2} be the signals transmitted at the source and relay nodes, then the received signals at relays and destination are given by

$$\begin{pmatrix} \mathbf{Y}_{r_1} \\ \mathbf{Y}_{r_2} \\ \mathbf{Y}_d \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 3 & 3 \end{pmatrix} \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_{r_1} \\ \mathbf{X}_{r_2} \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_{r_1} \\ \mathbf{Z}_{r_2} \\ \mathbf{Z}_d \end{pmatrix}$$

where $\mathbf{Z}_{r_1}, \mathbf{Z}_{r_2}, \mathbf{Z}_d$ are i.i.d. $\mathcal{N}(0, 1)$. For this example, we set the probability distribution of $\mathbf{X}_s, \mathbf{X}_{r_1}, \mathbf{X}_{r_2}$ to be jointly Gaussian with zero mean and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}.$$

Finally, consider the sets $A = \{s, r_1\}$ and $B = \{s, r_2\}$. Figure 1 shows how the function $F(A) + F(B) - F(A \cup B) - F(A \cap B)$ varies for different values of the correlation coefficient (between \mathbf{X}_{r_1} and \mathbf{X}_{r_2}) $\rho \in [0, 1]$. We see that $F(A) + F(B)$ can be greater than or less than $F(A \cup B) + F(A \cap B)$ depending on the value of ρ . It follows that in general $F(\cdot)$ is not a submodular or a supermodular function when there is correlation among the signals at the nodes.

IV. WIRELESS NETWORK MODELS

In this section, by applying the result of Theorem 1, we show that the capacity or an approximation to the capacity for the ADT deterministic network, Gaussian relay network, and wireless erasure network can be cast as a minimization of a submodular function.

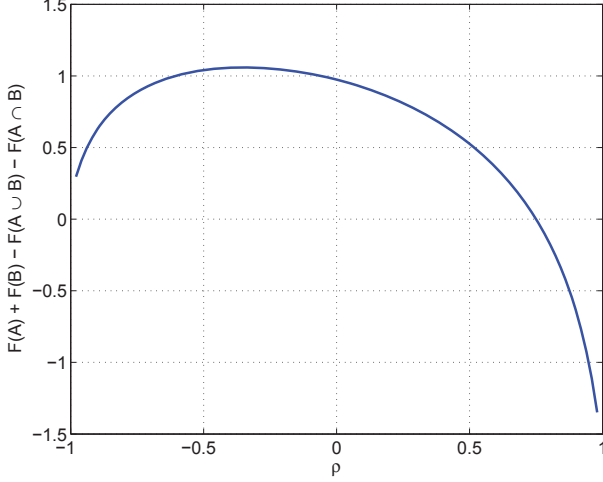


Fig. 1. $F(A) + F(B) - F(A \cup B) - F(A \cap B)$ as a function of the correlation coefficient ρ . In general, $F(\cdot)$ is neither submodular or supermodular when the signals are correlated.

A. Deterministic model (ADT)

We start by briefly describing the network model of [4]⁴. In this model, each link from node i to node j has an associated non-negative integer gain n_{ij} . Each node $i \in \mathcal{V}$ transmits a signal \mathbf{X}_i and receives a signal \mathbf{Y}_i , both in \mathbb{F}_p^q where $q = \max_{i,j} n_{ij}$. At any given time, the received signal at node j is given by

$$\mathbf{Y}_j = \sum_{i \in \mathcal{V} \setminus \{d\}} \mathbf{S}^{q-n_{ij}} \mathbf{X}_i \quad (1)$$

where d is the destination node, the shifting matrix \mathbf{S} is given by

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

and the sums and products are in \mathbb{F}_p .

For a given cut Ω of the network, where Ω includes the source node and excludes the destination node, we can stack together the input vectors $\mathbf{X}_i, i \in \Omega$ and output vectors $\mathbf{Y}_i, i \in \Omega^c$, and define a transition matrix Λ_Ω that gives the input-output relationship of these vectors according to (1). It is shown in [2] that the capacity of the deterministic network is equal to $\min_\Omega \text{rank}(\Lambda_\Omega)$. We show next in Theorem 2 that $\text{rank}(\Lambda_\Omega)$ is submodular, and hence the capacity can be computed by minimizing a submodular function.

Proposition 1. Assume an $m \times n$ matrix A over \mathbb{F}_p . Let \mathcal{N} be the subspace $\mathcal{N} \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{F}_p^n \mid A\mathbf{x} = 0\}$, and let \mathcal{G} be the set of cosets of \mathcal{N} in \mathbb{F}_p^n . Pick $\hat{\mathbf{x}}_i$ to be an element in the i th coset of \mathcal{N} for $i = 1, 2, \dots, |\mathcal{G}|$, and set $\mathbf{y}_i = A\hat{\mathbf{x}}_i$. Notice that $\mathbf{y}_i \neq \mathbf{y}_j$ if $i \neq j$. Now, if we choose \mathbf{x} uniformly at random from elements of \mathbb{F}_p^n with probability $1/|\mathbb{F}_p^n|$, then the mapping

$A\mathbf{x}$ maps \mathbf{x} to $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{|\mathcal{G}|}\}$ uniformly at random with probability $1/|\mathcal{G}|$. In addition, the cosets of \mathcal{N} form a partition of \mathbb{F}_p^n into $p^n/|\mathcal{N}|$ sets. Also $\text{rank}(A) + \log_p(|\mathcal{N}|) = n$. Thus, $\log_p |\mathcal{G}| = \text{rank}(A)$.

Theorem 2. For a deterministic model, given a cut Ω assume Λ_Ω is the transition matrix from nodes in Ω to nodes in Ω^c . Set $D(\Omega) = \text{rank}(\Lambda_\Omega)$, then $D(\Omega)$ is submodular.

Remark 1. A special case of Theorem 2 for layered ADT networks was proved in earlier works [13], [23].

Proof: In the network, assume node i sends b_i symbols $x_{i,1}, x_{i,2}, \dots, x_{i,b_i}$ with $x_{i,j} \in \mathbb{F}_p$. We assume $x_{i,j}$'s drawn i.i.d. with uniform probability distribution over \mathbb{F}_p , i.e. $p(x_{i,j} = q) = 1/|\mathbb{F}_p|$ for all $q \in \mathbb{F}_p$. From the definition of transition matrix, Λ_Ω , if we assume for the cut Ω , $\mathbf{s} = (s_1, s_2, \dots, s_k)^t$ symbols are being sent from nodes in Ω and $\mathbf{r} = (r_1, r_2, \dots, r_\ell)^t$ symbols are being received by nodes in Ω^c then $\mathbf{r} = \Lambda_\Omega \mathbf{s}$. Then we can write

$$\begin{aligned} I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c}) &= H(\mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c}) - H(\mathbf{Y}_{\Omega^c} | \mathbf{X}_\Omega, \mathbf{X}_{\Omega^c}) \\ &\stackrel{(a)}{=} H(\mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c}) \\ &= H(\Lambda_\Omega \mathbf{s} | \mathbf{X}_{\Omega^c}) \\ &\stackrel{(b)}{=} \log_p |\mathcal{G}| = \text{rank}(\Lambda_\Omega) \end{aligned}$$

where \mathcal{G} is the set of cosets of \mathcal{N} where $\mathcal{N} = \{\mathbf{s} : \Lambda_\Omega \mathbf{s} = 0\}$. Equality (a) is because \mathbf{Y}_{Ω^c} is a deterministic function of \mathbf{X}_Ω and (b) is the result of Proposition 1 and the fact the \mathbf{s} has uniform probability distribution.

Notice that for the independent probability distribution on the sources the received signals are independent conditioned on transmitted signals so, based on Theorem 1, $I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c})$ which is equal to $D(\Omega)$ is submodular. ■

B. Gaussian relay network

The Gaussian network model captures the effects of broadcasting, superposition and noise of power constrained wireless networks. In this model, at any time index (which we omit) the received signal at node $j \in \mathcal{V} \setminus \{s\}$ is given by

$$Y_j = \sum_{i \in \mathcal{V} \setminus \{d\}} h_{ij} X_i + N_j \quad (2)$$

where $X_i \in \mathbb{C}$ is the transmitted signal at node i , subject to an average power constraint $E(|X_i|^2) \leq 1$, $h_{ij} \in \mathbb{C}$ is the channel gain from node i to node j , and $N_j \in \mathcal{CN}(0, 1)$ is additive white circularly symmetric complex Gaussian noise, independent for different j .

It has been show in [21, Theorem 2.1] that using lattice codes for transmission and quantization at the relays, all rates R between source $\{s\}$ and destination $\{d\}$ satisfying

$$R \leq \min_\Omega I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c}) - |\mathcal{V}| \quad (3)$$

can be achieved, where Ω is a source-destination cut of the network and $\mathbf{X}_\Omega = \{X_i, i \in \Omega\}$ are i.i.d. $\mathcal{CN}(0, 1)$. In addition, the restriction to i.i.d. Gaussian input distributions is within $|\mathcal{V}|$ bits/s/Hz of the cut-set upper bound [2]. Therefore

⁴Please, refer [4] for a more complete description of the model and its motivation.

the rate achieved using lattice codes in the above result is within $2|\mathcal{V}|$ bits/s/Hz of the capacity of the network.

The following corollary is an immediate consequence of Theorem 1.

Corollary 1. *The function $F(\Omega) = I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c})$ with the elements of \mathbf{X}_Ω being i.i.d. $\mathcal{CN}(0, 1)$ is submodular.*

Due to Corollary 1 the minimization in (3) is the minimization of a submodular function and the resulting optimal value is within $2|\mathcal{V}|$ of the capacity of the network.⁵

C. Wireless erasure network

In [8] the authors introduce a special class of wireless networks, called wireless erasure networks. In these networks, a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ defines the interconnections between nodes. To model the broadcast effect of wireless networks, the signals on all outgoing arcs of any given node are equal to each other. There is no interference among multiple arcs arriving at a given node in this model, and the signals on the various arcs are erased independently of each other. We assume binary transmitted signals at each node, i.e. $X_i \in \{0, 1\}$, $i \in \mathcal{V} \setminus \{d\}$, but all the results can be extended to models with larger input alphabets. It has been shown in [8] that the capacity of the network is

$$C = \min_{\Omega} F(\Omega) = \min_{\Omega} \sum_{i \in \Omega} \left(1 - \prod_{j \in \Omega^c} \epsilon_{ij} \right) \quad (4)$$

where ϵ_{ij} is the probability of erasure when node i is sending information to node j . We show in the following theorem that $F(\Omega)$ is submodular.

Theorem 3. *The function $F(\Omega) = \sum_{i \in \Omega} \left(1 - \prod_{j \in \Omega^c} \epsilon_{ij} \right)$ equals $I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c})$ where X_i are i.i.d. $\sim \text{Bernoulli}(1/2)$ for $i \in \Omega$. Therefore, $F(\Omega)$ is submodular.*

Proof: For i.i.d. $X_i \sim \text{Bernoulli}(1/2)$, we can write

$$\begin{aligned} I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c}) &= H(\mathbf{X}_\Omega | \mathbf{X}_{\Omega^c}) - H(\mathbf{X}_\Omega | \mathbf{Y}_{\Omega^c}, \mathbf{X}_{\Omega^c}) \\ &\stackrel{(a)}{=} \sum_{i \in \Omega} (H(X_i) - H(X_i | \mathbf{Y}_{\Omega^c})) \\ &\stackrel{(b)}{=} \sum_{i \in \Omega} (1 - H(X_i | \mathbf{Y}_{\Omega^c})) \\ &= \sum_{i \in \Omega} \left(1 - \sum_{y_j \in \{1, 0, e\}, j \in \Omega^c} H(X_i | Y_j = y_j, j \in \Omega^c) p(Y_j = y_j, j \in \Omega^c) \right) \\ &= \sum_{i \in \Omega} \left(1 - H(X_i | Y_j = e, j \in \Omega^c) p(Y_j = e, j \in \Omega^c) \right) \\ &\stackrel{(c)}{=} \sum_{i \in \Omega} \left(1 - \prod_{j \in \Omega^c} \epsilon_{ij} \right). \end{aligned}$$

⁵Notice that $I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c}) = \log \det(I + HH^\dagger)$ where H is the matrix of channel gains from nodes in Ω to nodes in Ω^c and H^\dagger is the conjugate transpose of H . Therefore, it is easy to compute the capacity of each cut.

We used in (a) the independence among X_i and the channel erasures, in (b) the fact that for $X_i \sim \text{Bernoulli}(1/2)$, $H(X_i) = 1$, and in (c) the fact that for $X_i \sim \text{Bernoulli}(1/2)$, $H(X_i | Y_j = e, j \in \Omega^c) = 1$ and for independent erasures we have $p(Y_j = e, j \in \Omega^c) = \prod_{j \in \Omega^c} \epsilon_{ij}$. Theorem 1 can be applied to conclude that $F(\Omega)$ is submodular. ■

V. POWER OPTIMIZATION

In the previous section, for the Gaussian relay network model, we considered *fixed power assignments* to the different nodes in the network, and have shown that a constant gap approximation to the capacity can be efficiently computed. In many applications it is of interest to allocate the nodes' transmission powers to optimize a given objective. For example, in a network where the nodes are battery powered, it may be of interest to maximize the network lifetime while satisfying a baseline quality of service. Alternatively, it may be desirable to maximize the network throughput for a given total power budget. This total power budget may arise due to, e.g., a maximum system weight constraint which is dominated by the battery weight, or a total system cost, which may be heavily influenced by the cost of the batteries. Power allocation optimization may also naturally arise in situations where the channel gains, while known to all the nodes, slowly vary over time. In this case, it may be desirable to optimally allocate power for the current channel condition.

As before, we characterize communication rates in terms of cut-set capacities. We consider a model where the cut-set capacities are functions of the cuts and powers assigned to the nodes in the network: $F(\Omega, \mathbf{p}) : \mathcal{V} \times \mathbb{R}^{|\mathcal{V}|} \rightarrow \mathbb{R}$, and we focus on the Gaussian model where this function depends explicitly on the power assignment,

$$F(\Omega, \mathbf{p}) = I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c} | \mathbf{X}_{\Omega^c}) = \log \det(I + H_\Omega P_\Omega H_\Omega^\dagger) \quad (5)$$

where H_Ω is the matrix of channel gains from nodes in Ω to nodes in Ω^c , H^\dagger is the conjugate transpose of H , and P_Ω is a diagonal matrix where the diagonal elements are the powers of the nodes in Ω .

We will show in Lemma 1 below that in the Gaussian case $F(\Omega, \mathbf{p})$ is a concave function of \mathbf{p} . While the results of this section are stated and proved for the Gaussian model, we conjecture that similar results should hold for other models in which $F(\Omega, \mathbf{p})$ is a concave function of \mathbf{p} .

For Gaussian relay networks, we show that the following optimization problem can be solved in polynomial time,

$$\begin{aligned} &\underset{R, \mathbf{p}, \mathbf{P}}{\text{minimize}} && \mu_1 R + \mu_2 P \\ &\text{subject to} && R \leq F(\Omega \cup \{s\}, \mathbf{p}) \text{ for all } \Omega \subseteq \mathcal{V} \setminus \{d\} \\ &&& 0 \leq \mathbf{p} \leq \mathbf{p}_{\max} \\ &&& \sum_{i=1}^{|\mathcal{V}|} p_i \leq P \\ &&& R_0 \leq R, P \leq P_{\text{tot}} \end{aligned} \quad (6)$$

for fixed constants $\mu_1, \mu_2, R_0, \mathbf{p}_{\max}$ and P_{tot} . In the rest of the section, we denote the feasible set of the optimization (6) by \mathcal{K} .

We use the Ellipsoid method [14], [15] to show that the optimization (6) can be solved efficiently. We will use the following definitions and result. The reader is referred to [15] for more details.

Definition 1 (Polynomial computability). *A family of optimization programs is polynomially computable if:*

- (i) *for any instance of the program and any point \mathbf{x} in the domain, the objective and its subgradient can be computed in polynomial time in the size of the instance.*
- (ii) *for a given measure of infeasibility $\text{Infeas}(\cdot)$, it should be possible to determine if $\text{Infeas}(\mathbf{x}) \leq \varepsilon$ in polynomial time, and when this inequality is not satisfied, it should be possible to find in polynomial time a vector \mathbf{c} such that*

$$\mathbf{c}^T \mathbf{x} > \mathbf{c}^T \mathbf{y}, \forall \mathbf{y} : \text{Infeas}(\mathbf{y}) \leq \varepsilon.$$

Definition 2 (Polynomial growth). *A family of optimization programs has polynomial growth if the objectives and the infeasibility measures as functions of points \mathbf{x} in the domain grow polynomially with $\|\mathbf{x}\|_1$.*

Definition 3 (Polynomial boundedness of feasible sets). *A family of optimization programs has polynomially bounded feasible sets if the feasible set of an instance of the program is contained in an Euclidean ball centered at the origin with radius that grows at most polynomially with the size of the instance.*

Proposition 2 ([15, Theorem 5.3.1]). *Let \mathcal{P} be a family of convex optimization programs equipped with infeasibility measure $\text{Infeas}(\cdot)$. Assume that the family is polynomially computable with polynomial growth and with polynomially bounded feasible sets. Then \mathcal{P} is polynomially solvable.*

In order to use Proposition 2 we need to check that the optimization (6) is a convex program. Since the objective function is linear, we only need to check that the feasible set \mathcal{K} is convex.

Lemma 1. *The feasible set \mathcal{K} is a convex set.*

Proof: First we show that the function $F(\Omega, \mathbf{p}) = \log \det(I + H_\Omega P_\Omega H_\Omega^\dagger)$ is concave in \mathbf{p} where $0 \leq \mathbf{p}$ for any cut $\Omega \subseteq \mathcal{V}$. For any two vectors $\mathbf{p}_1, \mathbf{p}_2 \geq 0$ and $\gamma \in [0, 1]$ we can write

$$\begin{aligned} & \gamma F(\Omega, \mathbf{p}_1) + (1 - \gamma) F(\Omega, \mathbf{p}_2) \\ &= \gamma \log \det(I + H_\Omega P_1 H_\Omega^\dagger) \\ & \quad + (1 - \gamma) \log \det(I + H_\Omega P_2 H_\Omega^\dagger) \\ & \stackrel{(a)}{\leq} \log \det(\gamma(I + H_\Omega P_1 H_\Omega^\dagger) \\ & \quad + (1 - \gamma)(I + H_\Omega P_2 H_\Omega^\dagger)) \\ &= \log \det(I + H_\Omega(\gamma P_1 + (1 - \gamma) P_2) H_\Omega^\dagger) \end{aligned}$$

where P_1 (P_2) is a diagonal matrix where the diagonal elements are the elements of \mathbf{p}_1 (\mathbf{p}_2) that belong to Ω (respectively), and (a) follows from the concavity of $\log \det X$, for $X \succ 0$ [24].

Next, consider the set $\mathcal{C}(\Omega) = \{(R, P, \mathbf{p}) : R \leq F(\Omega, \mathbf{p})\}$. Choose two vectors (R_1, P_1, \mathbf{p}_1) and (R_2, P_2, \mathbf{p}_2) in $\mathcal{C}(\Omega)$.

We will show that $\mathcal{C}(\Omega)$ is a convex set by showing that for any $\gamma \in [0, 1]$ the vector $\gamma(R_1, P_1, \mathbf{p}_1) + (1 - \gamma)(R_2, P_2, \mathbf{p}_2)$ is also in $\mathcal{C}(\Omega)$. Notice that $R_1 \leq F(\Omega, \mathbf{p}_1)$ and $R_2 \leq F(\Omega, \mathbf{p}_2)$ if and only if $(R_1, P_1, \mathbf{p}_1) \in \mathcal{C}(\Omega)$ and $(R_2, P_2, \mathbf{p}_2) \in \mathcal{C}(\Omega)$. Therefore

$$\begin{aligned} & \gamma R_1 + (1 - \gamma) R_2 \leq \gamma F(\Omega, \mathbf{p}_1) + (1 - \gamma) F(\Omega, \mathbf{p}_2) \\ & \stackrel{(a)}{\leq} F(\Omega, \gamma \mathbf{p}_1 + (1 - \gamma) \mathbf{p}_2) \end{aligned}$$

where (a) is due to the fact that $F(\Omega, \mathbf{p})$ is a concave function with respect to \mathbf{p} . Thus, $\mathcal{C}(\Omega)$ is a convex set for any $\Omega \subseteq \mathcal{V}$.

It is easy to check that the sets $\mathcal{P}_1 = \{(R, P, \mathbf{p}) : 0 \leq \mathbf{p} \leq \mathbf{p}_{\max}\}$, $\mathcal{P}_2 = \{(R, P, \mathbf{p}) : \sum p_i \leq P\}$, $\mathcal{P}_3 = \{(R, P, \mathbf{p}) : R_0 \leq R\}$, and $\mathcal{P}_4 = \{(R, P, \mathbf{p}) : P \leq P_{\text{tot}}\}$ are also convex sets, and, as a result, $\mathcal{K} = \bigcap_{\Omega \subseteq \mathcal{V} \setminus \{d\}} \mathcal{C}(\Omega \cup \{s\}) \cap \mathcal{P}_1 \cap \mathcal{P}_2 \cap \mathcal{P}_3 \cap \mathcal{P}_4$ is a convex set. ■

Having proved that (6) is a convex program, in order to use Proposition 2 we need to check that the conditions of Definitions 1, 2, and 3 are satisfied. Part (i) of Definition 1 follows from the linearity of the objective in (6). For part (ii) of Definition 1 we specify an infeasibility measure $\text{Infeas}(\cdot) : \mathbb{R}^{|\mathcal{V}|+2} \rightarrow \mathbb{R}$ as follows⁶:

$$\begin{aligned} \text{Infeas}((R, P, \mathbf{p})) = \max \left\{ 0, -\mathbf{p}, \mathbf{p} - \mathbf{p}_{\max}, R_0 - R, P - P_{\text{tot}} \right. \\ \left. \sum_{i=1}^{|\mathcal{V}|} p_i - P, R - \min_{\Omega \in \mathcal{V} \setminus \{d\}} F(\Omega \cup \{s\}, \mathbf{p}) \right\} \end{aligned} \quad (7)$$

The conditions of part (ii) of Definition 1 are verified in the following theorem.

Theorem 4. *For a given vector $(R, P, \mathbf{p}) \in \mathbb{R}^{|\mathcal{V}|+2}$ and any $\varepsilon > 0$ we can either (a) determine in polynomial time if $\text{Infeas}((R, P, \mathbf{p})) \leq \varepsilon$ and if not (b) find in polynomial time a vector $\mathbf{c} \in \mathbb{R}^{|\mathcal{V}|+2}$, such that for every (R', P', \mathbf{p}') satisfying $\text{Infeas}((R', P', \mathbf{p}')) \leq \varepsilon$, $\mathbf{c}^T (R', P', \mathbf{p}') < \mathbf{c}^T (R, P, \mathbf{p})$.*

Proof: Part (a) requires checking that each of the arguments of the max of (7) is smaller than or equal to ε in polynomial time. The first six terms are linear functions and can be easily computed. The last term can be compared to ε by performing a minimization of a submodular function, which as was shown in Section IV-B, can also be computed in polynomial time.

We focus on condition (b). In this case $\text{Infeas}((R, P, \mathbf{p})) > \varepsilon$, meaning that at least one of the arguments of the max of (7) is larger than ε . We consider each case separately.

If $-p_i > \varepsilon$ we set $\mathbf{c} = -\mathbf{e}_{i+2}$ where \mathbf{e}_i has a one in the i^{th} position and zeros everywhere else, which can be easily checked to satisfy the condition of part (b).

Similarly, for the cases $p_i - p_{\max, i} > \varepsilon$, $R_0 - R > \varepsilon$, $P - P_{\text{tot}} > \varepsilon$ and $\sum_{i=1}^{|\mathcal{V}|} p_i - P$ we set $\mathbf{c} = \mathbf{e}_{i+2}$, $\mathbf{c} = -\mathbf{e}_1$, $\mathbf{c} = \mathbf{e}_2$, $\mathbf{c} = (0, -1, 1, \dots, 1)$ respectively.

For the last case, let $\Omega^* = \arg \min_{\Omega \in \mathcal{V} \setminus \{d\}} F(\Omega \cup \{s\}, \mathbf{p})$. We have $R - F(\Omega^* \cup \{s\}, \mathbf{p}) > \varepsilon$. Since the function $F(\Omega^*, \mathbf{p})$

⁶With a slight abuse of notation we use max on vector quantities by taking the maximum over the components of the vector.

is continuous and differentiable with respect to \mathbf{p} , and the set $\tilde{\mathcal{C}}(\Omega^*, \varepsilon) = \{(R, P, \mathbf{p}) : R \leq F(\Omega^*, \mathbf{p}) + \varepsilon\}$ is convex (which can be shown as in the proof of Lemma 1), then the vector $\mathbf{c} = (1, 0, -\nabla_{\mathbf{p}} F(\Omega^*, \mathbf{p}))$ is the normal to a hyperplane that separates (R, P, \mathbf{p}) from the set $\tilde{\mathcal{C}}(\Omega^*, \varepsilon)$. In other words, for all $(R', P', \mathbf{p}') \in \mathcal{C}(\Omega^*)$ we have $\mathbf{c}^T(R', P', \mathbf{p}') < \mathbf{c}^T(R, P, \mathbf{p})$. Noting that $\{(R', P', \mathbf{p}') : \text{Infeas}((R', P', \mathbf{p}')) \leq \varepsilon\} \subseteq \tilde{\mathcal{C}}(\Omega^*, \varepsilon)$, we conclude that part (b) holds in this case as well. ■

Having proved these preliminary results, we are ready to prove the main result of this section.

Theorem 5. *The optimization in (6) can be solved in polynomial time on the size of the problem.*

Proof: The proof uses Proposition 2, which requires verifying the convexity of the problem together with the conditions of polynomial computability, polynomial growth, and polynomial boundedness of the feasible set. Convexity was proved in Lemma 1, while polynomial computability was shown in Theorem 4. Polynomial growth follows from the fact that $F(\Omega, \mathbf{p})$ is the log of a polynomial on \mathbf{p} with degree at most $|\mathcal{V}|$, while the objective and remaining terms that define the infeasibility measure are linear on (R, P, \mathbf{p}) . Finally, to check that feasible set is polynomially bounded, we note that the feasible set is a subset of the hypercube

$$\{(R, P, \mathbf{p}) : 0 \leq (R, P, \mathbf{p}) \leq (R_{\max}, P_{\text{tot}}, \mathbf{p}_{\max})\}$$

where $R_{\max} = \min_{\Omega \in \mathcal{V} \setminus \{d\}} F(\Omega \cup \{s\}, \mathbf{p}_{\max})$. It follows that the feasible set is contained in the Euclidean ball centered at the origin with radius $\|(R_{\max}, P_{\text{tot}}, \mathbf{p}_{\max})\|_2$, which can be easily checked to grow polynomially on the size of the problem. ■

The general optimization problem (6) can be specialized to a power minimization with a minimum rate constraint, and to a rate maximization with a total power constraint. Both problems can be solved in polynomial time, as stated in the following corollaries to Theorem 5.

Corollary 2. *The following power minimization problem can be solved in polynomial time.*

$$\begin{aligned} & \underset{\mathbf{p}}{\text{minimize}} && \sum_{i=1}^{|\mathcal{V}|} p_i \\ & \text{subject to} && R_0 \leq F(\Omega \cup \{s\}, \mathbf{p}) \text{ for all } \Omega \subseteq \mathcal{V} \setminus \{d\} \\ & && 0 \leq \mathbf{p} \leq \mathbf{p}_{\max}. \end{aligned} \quad (8)$$

Proof: The corollary follows from Theorem 5 by setting $\mu_1 = 0$, $\mu_2 = 1$, $P_{\text{tot}} = \sum_{i=1}^{|\mathcal{V}|} p_{\max, i}$. ■

Corollary 3. *The following rate maximization can be solved in polynomial time.*

$$\begin{aligned} & \underset{R, \mathbf{p}}{\text{maximize}} && R \\ & \text{subject to} && 0 \leq R \leq F(\Omega \cup \{s\}, \mathbf{p}) \text{ for all } \Omega \subseteq \mathcal{V} \setminus \{d\} \\ & && \sum_{i=1}^{|\mathcal{V}|} p_i \leq P_{\text{tot}} \\ & && 0 \leq \mathbf{p} \leq \mathbf{p}_{\max} \end{aligned} \quad (9)$$

Proof: The corollary follows from Theorem 5 by setting $\mu_1 = -1$, $\mu_2 = 0$ and $R_0 = 0$. ■

VI. ALGORITHMS AND SIMULATIONS

In this section we study different algorithms and provide simulation results regarding submodular function minimization and power allocation problems. In the first subsection we look at minimum norm algorithm for submodular function minimization and we show that this algorithm can find the approximation to capacity of layered Gaussian relay networks with more than 300 nodes in a couple of minutes. In the second subsection we propose a heuristic algorithm to find the optimum power allocation for Gaussian relay networks.

A. submodular function minimization

One approach to solve the submodular minimization problem due to Lovász is based on *extension* of the set function $f : 2^{\mathcal{V}} \rightarrow \mathbb{R}$ to a convex function $g : [0, 1]^{|\mathcal{V}|} \rightarrow \mathbb{R}$ that agrees with f on the vertices of the hypercube $[0, 1]^{|\mathcal{V}|}$, with a guarantee that $\min_{A \subseteq \mathcal{V}} f(A)$ is equal to $\min_{\mathbf{x}} g(\mathbf{x})$ for $\mathbf{x} \in [0, 1]^{|\mathcal{V}|}$. In this section we assume the normalization $f(\emptyset) = 0$.

The Lovász extension g of any set function f can be defined as follows. For a given $\mathbf{x} \in [0, 1]^{|\mathcal{V}|}$ order the elements of \mathcal{V} such that $x(v_1) \geq x(v_2) \geq \dots \geq x(v_n)$, where $x(v_i)$ is the v_i th element of the vector \mathbf{x} . Set $\lambda_0 = 1 - x(v_1)$, $\lambda_i = x(v_i) - x(v_{i+1})$, $\lambda_n = x(v_n)$, and

$$g(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i=1}^n \lambda_i f(\{v_1, v_2, \dots, v_i\}).$$

Define $\mathbf{1}_{\emptyset} = \mathbf{0} \in \mathbb{R}^n$ and $\mathbf{1}_{\{v_1, v_2, \dots, v_i\}}$ as an n dimensional vector such that the coordinates v_1, v_2, \dots, v_i are equal to one and all the other coordinates are equal to zero. Then, it is easy to see that $\mathbf{x} = \sum_{i=0}^n \lambda_i \mathbf{1}_{\{v_1, v_2, \dots, v_i\}}$, $\sum_{i=0}^n \lambda_i = 1$ and $\lambda_i \geq 0$. So, \mathbf{x} is a unique linear convex combination of some vertices of the hypercube and $g(\mathbf{x})$ is linear convex combination of values of f on those vertices.

A key result is that if f is submodular its Lovász extension g is a convex function [14], [11]. In addition, finding the minimum of the submodular function f over subsets of \mathcal{V} is equivalent to finding the minimum of the convex function g in the hypercube $[0, 1]^{|\mathcal{V}|}$. The optimization can be done in polynomial time using Ellipsoid algorithm [14].

There are other algorithms with faster running time to solve the submodular minimization problem [16], [17], [20]. To the best of our knowledge, the running time of the fastest algorithm is in the order of $O(n^5 \alpha + n^6)$, where α is the time that the algorithms takes to compute $f(A)$ for any subset $A \subseteq \mathcal{V}$ [20]. For ADT networks, Gaussian relay networks, and erasure networks, α is the time to compute: the rank of $n \times n$ matrices, the determinant of $n \times n$ matrices, and equation (4), respectively.

However, for networks of large size, a complexity of $O(n^5 \alpha + n^6)$ may still be computationally cumbersome. As a result, in these cases it is desirable to have faster algorithms. Recently, Fujishima [11], [12] showed that the minimization of any submodular function can be cast as a

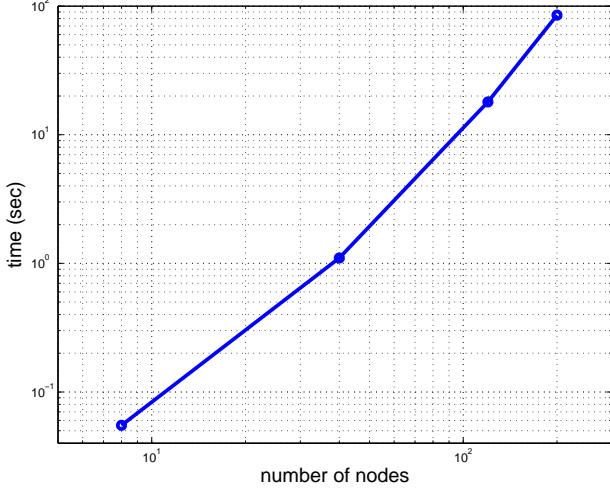


Fig. 2. Running time of minimum norm algorithm for a layered Gaussian relay network. Each layer has four nodes.

minimum norm optimization over the base polytope of f , $B_f = P_f \cap \{\mathbf{x} \mid \sum_{i \in \mathcal{V}} x(i) = f(\mathcal{V})\}$, where

$$P_f \stackrel{\text{def}}{=} \left\{ \mathbf{x} \in \mathbb{R}^n \mid \forall A \subseteq \mathcal{V} : \sum_{i \in A} x(i) \leq f(A) \right\}$$

and the corresponding minimum norm optimization is

$$\text{minimize } \|\mathbf{x}\|_2, \quad \text{subject to } \mathbf{x} \in B_f. \quad (10)$$

Letting \mathbf{x}^* be the solution of this minimization, the set $A^* = \{v_i : x^*(v_i) < 0\}$ is the solution to $\min_A f(A)$. Whether the above optimization problem can be solved in polynomial time is an open problem. However empirical studies [12] have shown that this algorithm has comparable or even faster running times than the other algorithms with polynomial time performance guarantees.

In our specific setting, for layered Gaussian relay networks of size up to around 300 nodes with 4 nodes per layer, we were able to find the approximate capacity (cf. (3)) in order of minutes (see Figure 2). In order to solve the minimization (10) we used the Matlab package provided in [19].

B. Power allocation

In Section V we have shown that the Ellipsoid method can be used to solve the optimization in (8) in polynomial time. While in theory this result shows that the optimization in (8) is tractable, in practice the Ellipsoid method has a number of shortcomings that limit its usability. On the one hand the running time of the Ellipsoid method can be large compared to alternative algorithms, and on the other hand, for high dimensional problems it has shown numerical instability. In this section, we propose a heuristic algorithm to solve the optimization in (8) and show that this algorithm converges to the right solution. While the running time can be exponential on the network size, we show through simulations that the algorithm often converges within a time proportional to the network size.

Our proposed algorithm, Algorithm 1 (see pseudo-code below), is based on the cutting plane methods [24] in convex optimization. The optimization in (8) contains exponentially many constraints of the form

$$R_0 \leq F(\Omega \cup \{s\}, \mathbf{p}) \text{ for } \Omega \subseteq \mathcal{V} \setminus \{d\}. \quad (11)$$

In Algorithm 1 we first find the min-cut corresponding to assigning maximum power to all nodes in the network:

$$\Omega_1 = \operatorname{argmin}_{\Omega \subseteq \mathcal{V} \setminus \{d\}} F(\Omega \cup \{s\}, \mathbf{p}_{\max}).$$

Then, we modify the optimization in (8) by replacing the constraint (11) with

$$R_0 \leq F(\Omega_1 \cup \{s\}, \mathbf{p}).$$

The resulting convex program can be easily solved since it contains few constraints.

After optimization, we let \mathbf{p}^* be the optimum power allocation for the current set of constraints and set

$$\Omega_i = \operatorname{argmin}_{\Omega \subseteq \mathcal{V} \setminus \{d\}} F(\Omega \cup \{s\}, \mathbf{p}^*).$$

We iteratively add the constraint

$$R_0 \leq F(\Omega_i \cup \{s\}, \mathbf{p})$$

to our set of constraints, and solve the optimization again. We stop if the new constraint is already in the set of constraints.

Algorithm 1 Power minimization

Input: Channel gain matrix H , desired rate R , vector of nodes' power constraints \mathbf{p}_{\max} .

Output: Min-cut, Power assignment \mathbf{p}^* that achieves approximate to rate R with minimum sum of powers.

```

 $\mathcal{C} \leftarrow \{\}, \mathbf{p}^* \leftarrow 0$ 
 $\Omega^* \leftarrow \min_{\Omega \subseteq \mathcal{V} \setminus \{d\}} F(\Omega \cup \{s\}, \mathbf{p}_{\max})$ 
if  $R \leq F(\Omega^* \cup \{s\}, \mathbf{p}_{\max})$  then
  while  $\Omega^* \notin \mathcal{C}$  and  $F(\Omega^* \cup \{s\}, \mathbf{p}^*) < R$  do
     $\mathcal{C} \leftarrow \mathcal{C} \cup \{\Omega^*\}$ 
     $\mathbf{p}^* \leftarrow \min \sum p_i$ 
    subject to:
       $R \leq F(\Omega \cup \{s\}, \mathbf{p})$  for all  $\Omega \in \mathcal{C}$ 
       $0 \leq \mathbf{p} \leq \mathbf{p}_{\max}$ 
     $\Omega^* \leftarrow \min_{\Omega \subseteq \mathcal{V} \setminus \{d\}} F(\Omega \cup \{s\}, \mathbf{p}^*)$ 
  end while
  return  $\Omega^* \cup \{s\}, \mathbf{p}^*$ 
else
  print The constraints are infeasible.
end if

```

Since in each iteration the algorithm adds a new constraint to the constraint set and the number of constraints in (11) is finite, the algorithm is guaranteed to find the optimum power in a finite number of iterations, which can be exponential.

We used simulations to test the performance of the algorithm for networks of varying size n ranging from 10 to 40. For each n , we generated 300 random networks with channels gains drawn i.i.d. using a $\mathcal{N}(0, 1)$ distribution. We set the desired transmission rate $R_0 = 4$ in (8) and set a maximum power constraint in each node to $p_{\max} = 100$. The results are

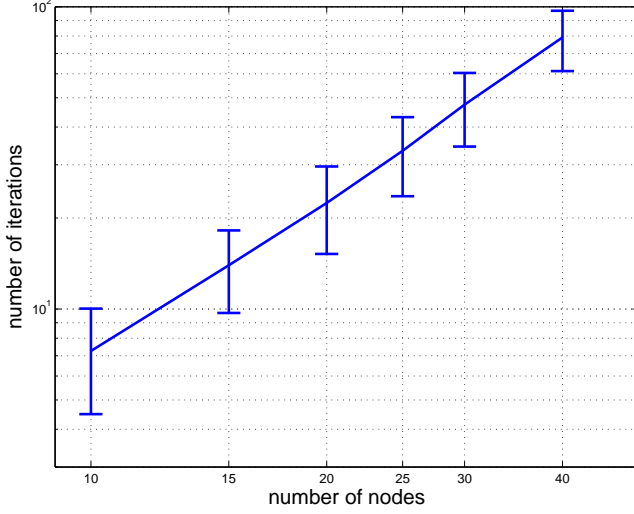


Fig. 3. Number of constraints in Algorithm 1 when the algorithm terminates. The error bars represent one standard deviation.

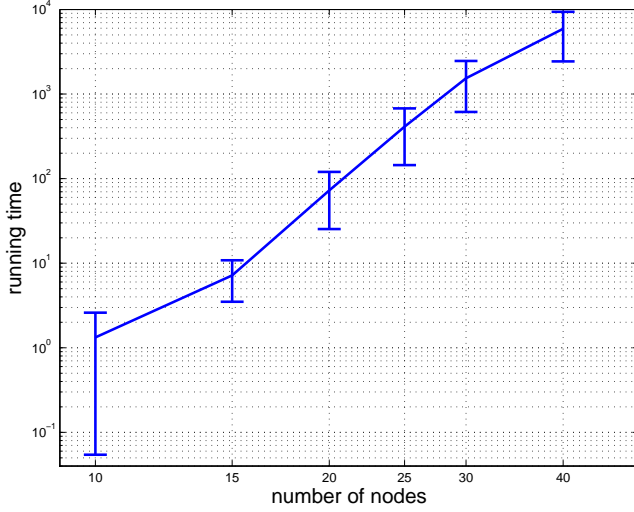


Fig. 4. Running time in seconds of the power optimization in Algorithm 1.

shown in Figures 3–6, where the the vertical bars represent ± 1 standard deviation around the mean computed over the 300 random networks.

Figure 3 shows the number of iterations of Algorithm 1 as a function of the number of nodes in the network. We see that the number of iterations grows $O(n^{3/2})$ with the number of nodes.

In Figure 4, we present the simulation time as a function of the network size. We observe a running time that grows slower than $O(n^6)$. The figure also shows that the power optimization of networks of 40 nodes completes in less than two hours on an Intel Xeon quad core CPU running at 2.33 GHz.

The minimum sum of powers to approximately achieve $R_0 = 4$ for networks of different size is presented in Figure 5 with a blue solid line. Interestingly, the plot shows that the minimum power concentrates around the mean. In addition, the figure shows diminishing returns in total power savings resulting from increasing the network size.

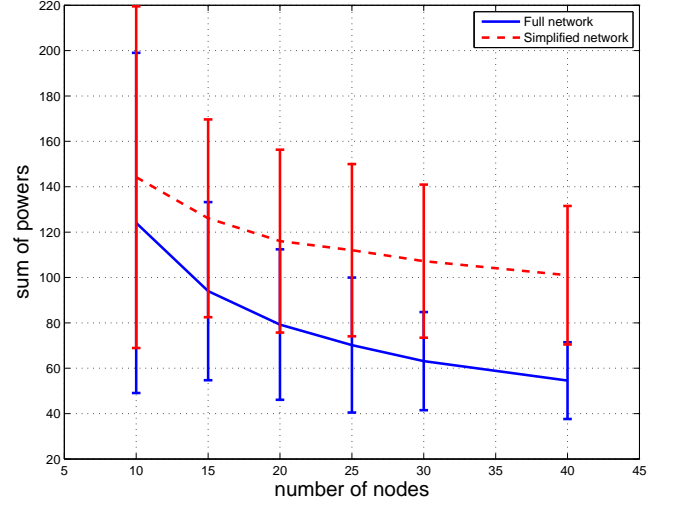


Fig. 5. Minimum sum of powers for optimization (8) when $R_0 = 4$ and the network is generated randomly as described in the paper.

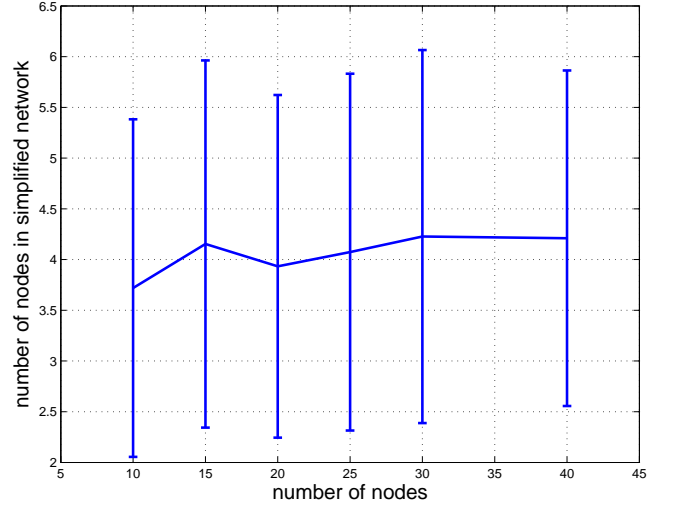


Fig. 6. Number of nodes in the simplified network.

For the special case of the diamond relay network with N relays, [26] shows that a fraction $k/(k+1)$ of the network capacity can be approximately achieved by using only k relays. In our setting of a general Gaussian relay network and sum-power minimization, we are interested in investigating whether it is possible to remove a large fraction of the nodes from the network without significantly affecting its performance.

In order to determine which nodes to remove from the network, we solve the sum-power minimization and compare the optimal power allocation p_i^* of each node i to a threshold P_{th} . All relays with $p_i^* < P_{th}$ are removed from the network. Let \mathcal{N} be the set of nodes with $p_i^* \geq P_{th}$. We optimize the power allocation for the network with node set $\{s\} \cup \mathcal{N} \cup \{d\}$ and determine whether the problem is feasible for $R_0 = 4$. If the problem is infeasible, we enlarge \mathcal{N} by adding more relay nodes in decreasing order of p_i^* , until the problem becomes feasible.

Figure 5 shows in the red dashed curve the resulting minimum sum of powers obtained by setting $P_{th} = 1$. Figure 6

shows the corresponding size of the set \mathcal{N} . We observe in Figure 6 that the number of nodes with allocated power p_i^* exceeding $P_{th} = 1$ (possibly including more relays to make the problem feasible) remains fairly constant as the size of the network n increases. This means that most of the power is allocated to a small subset of the nodes.

Removing the remaining nodes from the network and optimizing the power allocation again over the resulting simplified network results in the minimum total power plotted in Figure 5. This figure shows that even though the number of nodes in the simplified network remains approximately constant as n increases, the total power required to approximately achieve $R_0 = 4$ decreases with n . This is due to the fact that larger n allows to choose the best relays for the simplified network. There is some performance loss in terms of total power due to network simplification but this loss may be compensated by power savings arising from turning off some of the relays. While we have not modeled the power consumption of the clocks, CPU and other subsystems required to keep a relay active, in practice they may become comparable to the power consumed by radio transmissions. This makes network simplification very useful in practice.

REFERENCES

- [1] A. Amduruz and C. Fragouli, "Combinatorial algorithms for wireless information flow," *SODA '09: Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2009.
- [2] A. S. Avestimehr, S. N. Diggavi and D. N. C. Tse, "Approximate Capacity of Gaussian Relay Networks," *ISIT '08: IEEE International Symposium on Information Theory*, pp. 474–478, July 2008.
- [3] A. S. Avestimehr, S. N. Diggavi and D. N. C. Tse, "Wireless Network Information Flow: A Deterministic Approach," *IEEE Trans. Info. Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.
- [4] S. Avestimehr, S.N. Diggavi and D. N. C. Tse, "A deterministic approach to wireless relay networks," *Forty-Fifth Allerton Conference*, Illinois, September 2007.
- [5] Brett E. Schein, *Distributed coordination in network information theory*, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.
- [6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.
- [7] T. M. Cover and A. El Gamal, "Capacity Theorems for the Relay Channel," *IEEE Trans. Info. Theory*, vol. 25, no. 5, Sept. 1979, pp. 572–584.
- [8] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, M. Effros, "Capacity of wireless erasure networks," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 789–804, March 2006.
- [9] J. Ebrahimi, C. Fragouli, "Combinatorial Algorithms for Wireless Information Flow," *under submission in ACM Transactions in Algorithms*, 2009. Also available at arXiv:0909.4808v1.
- [10] E. Erez, Y. Xu and E. M. Yeh, "Coding for the Deterministic Network Model," *ITA workshop*, San Diego, 2010.
- [11] S. Fujishige, *Submodular functions and optimization*, (Second Edition), Annals of Discrete Mathematics, Vol. 58, Elsevier, Piscataway, N.J., U.S.A., 2005.
- [12] S. Fujishige, T. Hayashi, and S. Isotani, *The Minimum-Norm-Point Algorithm Applied to Submodular Function Minimization and Linear Programming*, Kyoto University, Kyoto, Japan, 2006.
- [13] M. X. Goemans, S. Iwata and R. Zenklusen, "An Algorithmic Framework for Wireless Information Flow," *Forty-Seventh Annual Allerton Conference*, October 2009.
- [14] M. Groetschel, L. Lovasz, and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization," *In Combinatorica*, 1:169–197, 1981.
- [15] A. Ben-Tal, and A. Nemirovski, "Lectures on Modern Convex Optimization. Analysis, Algorithms, and Engineering Applications," SIAM, Philadelphia, PA, 2001.
- [16] S. Iwata, "A faster scaling algorithm for minimizing submodular functions," *SIAM Journal on Computing*, 32(4):833–840, 2003.
- [17] S. Iwata and J. B. Orlin, "A simple combinatorial algorithm for submodular function minimization," In *SODA '09*, pp. 1230–1237, 2009.
- [18] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, 51(9):3037–3063, September 2005.
- [19] R. A. Krause, *Matlab Toolbox for Submodular Function Optimization*, <http://www.cs.caltech.edu/~krause/sfo/>
- [20] J. Orlin, "A faster strongly polynomial time algorithm for submodular function minimization," *Mathematical Programming*, 118(2): 237–251, 2009.
- [21] A. Özgür, and S. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice codes," *to be submitted to IEEE Journal on Selected Areas in Communications*, 2010. Available on-line at arXiv:1005.1284v1.
- [22] C. Shi and A. Ramamoorthy, "Improved combinatorial algorithms for wireless information flow," *Forty-Eighth Annual Allerton Conference*, pp. 875–880, October 2010.
- [23] S. M. Sadeh Tabatabaei Yazdi and S. A. Savari, "A combinatorial study of linear deterministic relay networks," *IEEE Information Theory Workshop (ITW)*, pp. 1–5, Jan 6–8, 2010.
- [24] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge university press, New York, 2004.
- [25] F. Parvaresh and R. Etkin, "On computing capacity of relay networks in polynomial time," *IEEE International Conference on Information Theory (ISIT)*, pp. 1342–1346, Saint Petersburg, Russia, August 2011.
- [26] C. Nazeroglu, A. Ozgur, J. Ebrahimi, and C. Fragouli, "Wireless Network Simplification: the Gaussian N-Relay Diamond Network with Multiple Antennas," *IEEE International Symposium on Information Theory (ISIT)*, Saint Petersburg, Russia, pp. 81–85, Aug. 2011.
- [27] A. Raja, and P. Viswanath, "Compress-and-Forward Scheme for a Relay Network: Approximate Optimality and Connection to Algebraic Flows," *IEEE International Conference on Information Theory (ISIT)*, pp. 1698–1702, Saint Petersburg, Russia, August 2011.